Tourte, G. J. L., Tonkin, E. L., Valdes, P. J., & Ball, A. (2010). *PEG-BOARD -- Open Access, Reuse and Preservation of Palæoclimate Data*. Poster session presented at 6th International Digital Curation Conference, Chicago, IL, United States.

PEG-BOARD - Open Access, Reuse and Preservation of Palæoclimate Data

Introduction

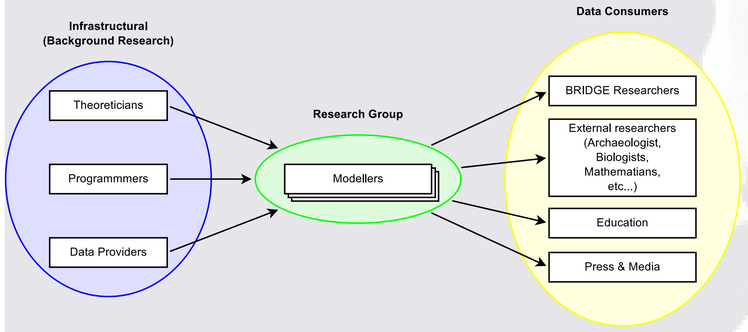
The BRIDGE research group was set up in 2003, and aims to improve the understanding of natural climate and environmental variability and to use this knowledge to predict future changes more accurately and assess its impact on all aspects of human society.

Data Preservation

The work on which the PEG-BOARD project focuses inherits from prior developments in the same area, including a software and service basis designed to support a primary activity of the research group - palæoclimate modeling via software simulation around sampled historical data. This service set is approximately fifteen years old. Palæoclimate simulations generate a great deal of data that is frequently reused by researchers in the sciences and humanities, worldwide. However, as a high-performance computing application, the quantities of data created are extremely large, meaning that participants have historically had to develop and apply de-facto preservation and data compression or summarisation policies. Since types of information required by users from areas as diverse as evolutionary biology, archæology and earth science vary greatly, the project has also developed,based on existing software, an in-house interface designed to support tailored information extraction from climate model information.

BRIDGE Data

The BRIDGE group has chosen to provide open access to data, with particular focus on those working internally to the research group, external researchers in the same field, and researchers with an interdisciplinary focus meshing in some way with the data available from BRIDGE palæoclimatology research, such as archæologists, biologists, and earth scientists. As user requirements have become evident over the past decade, a web site has been developed in an ad hoc manner that has enabled non-expert users to make use of a number of common methods of individual or comparative analysis, that is, to make use of the data for their own purposes.



Educational Reuse

The proportion of internal and external data reuse is approximately equal. Educational reuse occurs periodically following the term-based structure. Another major form of reuse is the research-oriented reuse, unsurprisingly since the platform operates primarily as an e-Science resource.

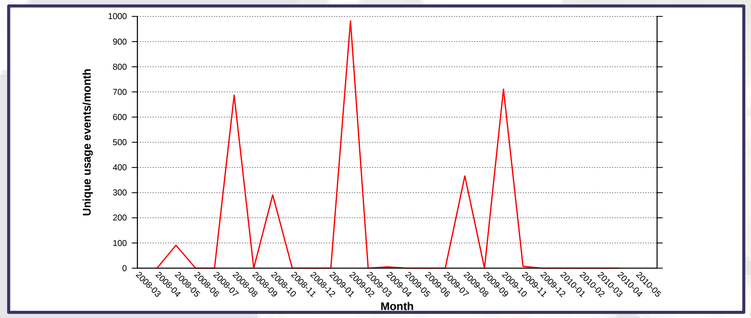
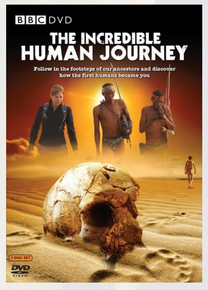


Figure 1: Example of educational reuse of a dataset over a 2 year period

Media Reuse

BBC's Incredible Human Journey is an example of media use of a dataset developed from a set of simulations originally designed for a single purpose. It is noticeable that it has been accessed a number of times, each time by different users. Whilst the first two spikes in reuse are likely to do with its initial use case, later accesses are pobably irrelevant, as the documentary had already been aired. There is a reasonably clean 'decay curve' visible in this graph, expected for resources with declining importance over time.



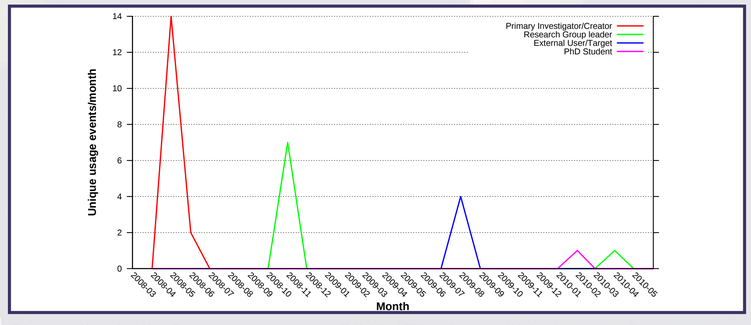


Figure 2: Media Reuse. Access to BBC's "The Incredible Journey" dataset

Discussion & Conclusion

This study shows us that each dataset beats with a different pulse, and that the key to gaining a better understanding of these patterns of use is to begin to model them according to the forces that drive each underlying process. To provide a detailed model for each form of reuse is perhaps beyond practical reach. Predictive modeling of reuse patterns of a given dataset shares a characteristic with climate data modelling: that is, that the behaviour of the wider system that drives interest and activity in a research domain might properly be described as having a chaotic element. Through data mining, we also begin to see how data sets interrelate.

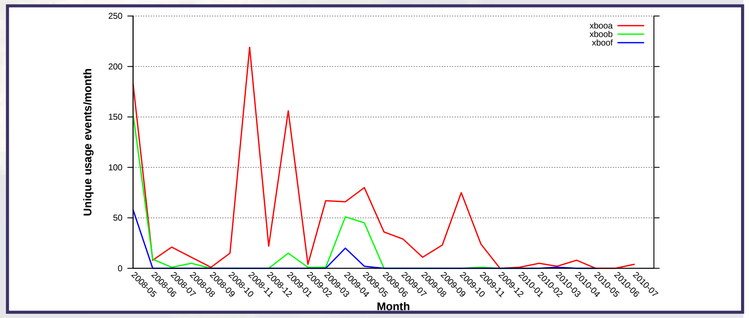


Figure 3: Example of usage decay curves

In future work, we will continue our development of data management policies and documentation, best practice and an enriched software backend as far as possible generic to any sort of climate data and as accessible as possible to most disciplines. We also intend to continue our research in data mining, seeking to improve our identification of usage patterns, and of relationships between datasets such as the obsolescence and replacement pattern.

Gregory J. L. Tourte
School of Geographical Sciences
The University of Bristol, UK
g.j.l.tourte@bristol.ac.uk

Prof. Paul Valdes
School of Geographical Sciences
The University of Bristol, UK
p.j.valdes@bristol.ac.uk

Emma L. Tonkin
UKOLN
The University of Bath, UK
e.tonkin@ukoln.ac.uk

Alex Ball
UKOLN
The University of Bath, UK
a.ball@ukoln.ac.uk

http://www.bridge.bristol.ac.uk/

http://www.paleo.bris.ac.uk/projects/peg-board/

References

[Journal article] Higgins, S. (2008). The DCC Curation Lifecycle Model, International Journal of Digital Curation, 3(1), 134-140.

[journal article] Jones, S., Ball, A., & Ekmekcioglu, C. (2008). The Data Audit Framework: A First Step in the Data Management Challenge. International Journal of Digital Curation, 3(2), 2008.

[report] ISO 14721:2003. Space data and information transfer systems -- Open archival information system -- Reference model. Geneva: International Organization for Standardization.

[proceedings] Tourte, G., Tonkin, E., & Valdes, P., 2010. (2010). The PEG-BOARD Project: A Case Study for BRIDGE. In: 14th International Conference on Electronic Publishing.